

UCSF

UC San Francisco Previously Published Works

Title

Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells.

Permalink

<https://escholarship.org/uc/item/5wh710z4>

Journal

BMC medicine, 9(1)

ISSN

1741-7015

Authors

Arasappan, Dhivya
Tong, Weida
Mummaneni, Padmaja
et al.

Publication Date

2011-05-01

DOI

10.1186/1741-7015-9-65

Peer reviewed

RESEARCH ARTICLE

Open Access

Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells

Dhivya Arasappan^{1,4}, Weida Tong^{1*}, Padmaja Mummaneni², Hong Fang³ and Shashi Amur^{2*}

Abstract

Background: A number of publications have reported the use of microarray technology to identify gene expression signatures to infer mechanisms and pathways associated with systemic lupus erythematosus (SLE) in human peripheral blood mononuclear cells. However, meta-analysis approaches with microarray data have not been well-explored in SLE.

Methods: In this study, a pathway-based meta-analysis was applied to four independent gene expression oligonucleotide microarray data sets to identify gene expression signatures for SLE, and these data sets were confirmed by a fifth independent data set.

Results: Differentially expressed genes (DEGs) were identified in each data set by comparing expression microarray data from control samples and SLE samples. Using Ingenuity Pathway Analysis software, pathways associated with the DEGs were identified in each of the four data sets. Using the leave one data set out pathway-based meta-analysis approach, a 37-gene metasignature was identified. This SLE metasignature clearly distinguished SLE patients from controls as observed by unsupervised learning methods. The final confirmation of the metasignature was achieved by applying the metasignature to a fifth independent data set.

Conclusions: The novel pathway-based meta-analysis approach proved to be a useful technique for grouping disparate microarray data sets. This technique allowed for validated conclusions to be drawn across four different data sets and confirmed by an independent fifth data set. The metasignature and pathways identified by using this approach may serve as a source for identifying therapeutic targets for SLE and may possibly be used for diagnostic and monitoring purposes. Moreover, the meta-analysis approach provides a simple, intuitive solution for combining disparate microarray data sets to identify a strong metasignature.

Please see Research Highlight: <http://genomemedicine.com/content/3/5/30>

* Correspondence: weida.tong@fda.hhs.gov; shashi.amur@fda.hhs.gov

¹National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA

²Genomics Group, Office of Clinical Pharmacology, Office of Translational Science, Center for Drug Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA
Full list of author information is available at the end of the article

Background

Microarrays are powerful tools with capability of measuring the transcript abundance of tens of thousands of genes simultaneously in biological samples. Microarray technology has matured over the past 15 years and is now employed for the study of gene expression signatures associated with disease [1-3]. The clinical utility of microarrays as prognostic tools can be evidenced by the approval of the US Food and Drug Administration (FDA) of a customized microarray, MammaPrint™ (Agendia, Amsterdam, The Netherlands) for predicting the outcomes in breast cancer patients on the basis of a 70-gene expression signature [4].

Some of the challenges associated with identification of gene expression signatures that differentiate the disease state from healthy controls are the availability of samples, sample size, heterogeneous data sets, and reproducibility. Thus, robustness of the gene expression signature derived from one study needs to be validated by other independent studies, preferably with large sample sizes. In practice, however, several studies with relatively small sample sizes are often used to identify gene expression signatures. In these circumstances, it is beneficial to combine the results of several individual studies using meta-analysis. This process enhances statistical power in identifying more robust and reliable gene signatures.

Several meta-analysis approaches have been proposed specifically for handling heterogeneous data sets. For example, Rhodes *et al.* [5] used the approach of utilizing *P* values of genes across studies to identify gene expression signatures that differentiate cancer tissues from normal tissues and to predict poor or good patient outcomes. Choi *et al.* [6] used an effect size estimate approach in a meta-analysis of two cDNA microarray data sets, human hepatocellular carcinoma and prostate cancer, to identify a transcriptional signature for cancer. A Bayesian approach was used by Wang *et al.* [7], who performed microarray studies on three different platforms and combined them to study differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells. Shen *et al.* [8] suggested a Bayesian mixture model incorporating the probability of expression measure.

Most of the currently used meta-analysis approaches first identify a set of commonly probed genes across studies and then derive a gene expression signature from these. A shortcoming of this approach is a potential loss of valuable information from individual data sets during the combining process. Thus, we propose a pathway-based meta-analysis approach whereby differentially expressed genes (DEGs) from individual studies are selected using a combination of *P* value and fold change

and the results are combined at the pathway level instead of at the gene level (see Figure 1 and Methods). Additionally, while most other methods perform very little validation or rely solely on the biological plausibility of the obtained results to serve as validation, the approach proposed here includes statistical validation through the leave one data set out permutation method. The results are further confirmed using an independent data set.

A number of authors have reported the use of microarray technology to identify gene expression signatures in systemic lupus erythematosus (SLE) [9-14], mechanisms underpinning SLE [15-17], and pathways related to SLE [18-20]. However, meta-analysis approaches have not been explored sufficiently in the study of SLE microarray data [21]. In the present study, the pathway-based meta-analysis method was applied to four independent gene expression oligonucleotide microarray data sets to identify gene expression signatures for SLE. These data sets were generated from peripheral blood mononuclear cell (PBMC) samples from SLE patients and healthy controls. The resulting signatures were then confirmed by testing on a fifth independent data set.

Methods

Data collection

Data sets from five independent microarray studies comparing PBMC samples from SLE patients with those from healthy individuals were obtained from prominent SLE researchers. These data sets are referred to as data sets 1, 2, 3, 4, and 5. Data sets 1, 2, 4, and 5 are associated with peer-reviewed publications [11-14]. Data set 3 is composed of unpublished data. Three of the studies (studies 1, 4, and 5) included only pediatric patients, while the remaining two included only adults. All studies employed the Affymetrix GeneChip microarray platform (Affymetrix, Inc., Santa Clara, CA, USA) but the versions of the array type varied (Table 1). In the case of two different array types used for the same study (that is, data sets 1 and 2), we treated them as separate data sets (data sets 1a, 1b, 2a, and 2b) during the meta-analysis. Raw data in the form of Affymetrix CEL files were provided for studies 1, 2, 3, and 5. For data set 4, however, expression values for a short-listed set of genes were provided. While data sets 1 to 4 were used in the meta-analysis workflow, data set 5 served as an independent data set to validate the gene signature derived from the meta-analysis.

Workflow of the pathway-based meta-analysis approach

The overall workflow of the pathway-based meta-analysis is summarized in Figure 1. The meta-analysis used a leave one data set out validation process. Both principal component analysis (PCA) and hierarchical cluster analysis

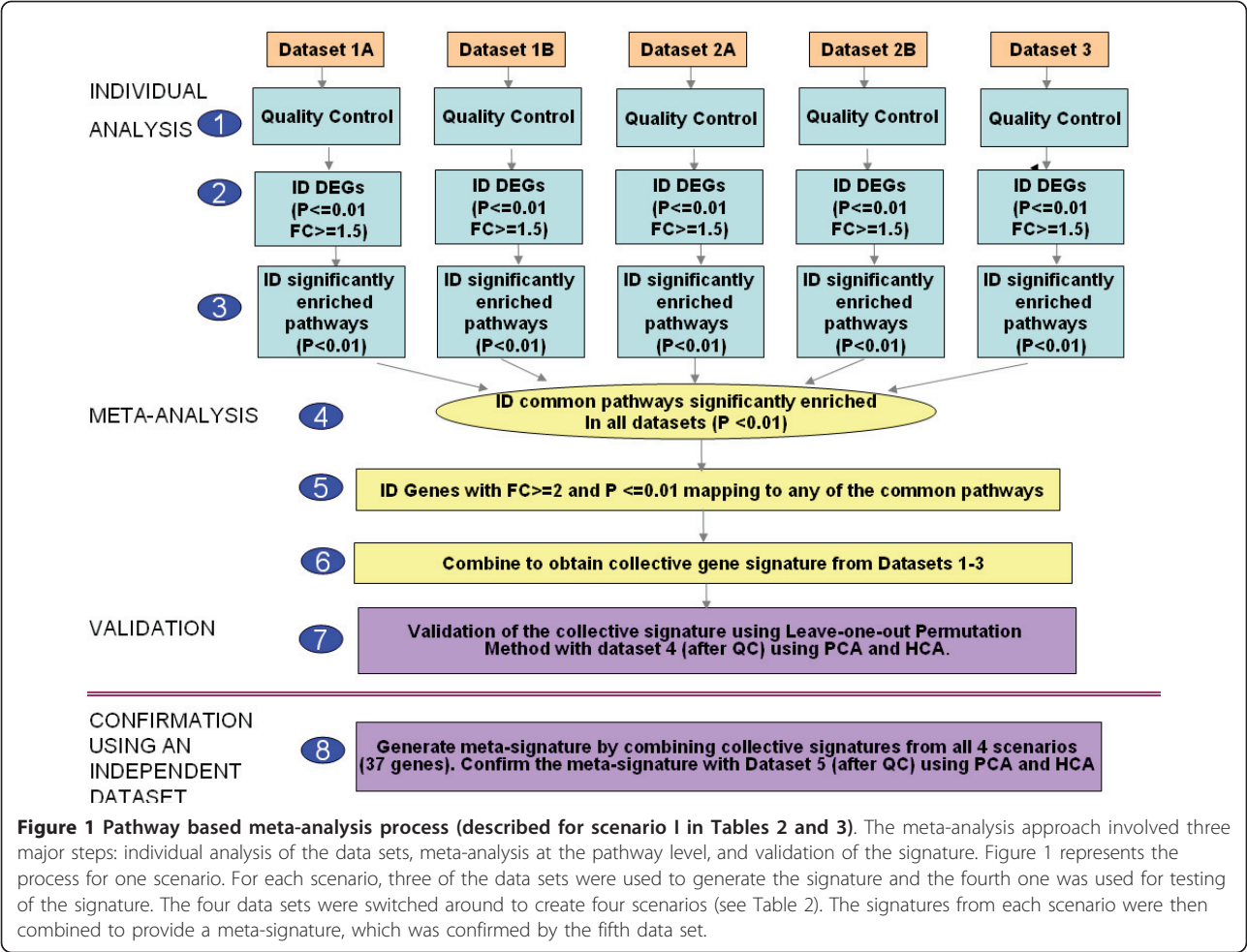


Figure 1 Pathway based meta-analysis process (described for scenario 1 in Tables 2 and 3). The meta-analysis approach involved three major steps: individual analysis of the data sets, meta-analysis at the pathway level, and validation of the signature. Figure 1 represents the process for one scenario. For each scenario, three of the data sets were used to generate the signature and the fourth one was used for testing of the signature. The four data sets were switched around to create four scenarios (see Table 2). The signatures from each scenario were then combined to provide a meta-signature, which was confirmed by the fifth data set.

(HCA) were used to visually inspect the leave one data set out cross-validation results. Last, the combined meta-signature obtained from the 4 data sets was validated against an independent fifth data set (data set 5).

For the individual quality control and data analysis steps mentioned below, each data set was considered separately. Additionally, since data sets 1 and 2 used two chip types each, they were considered as four different data sets (1a, 1b, 2a, and 2b) for the initial analysis.

Quality control

Quality assessment was done for each data set using the Genedata Expressionist (Genedata, San Francisco, CA,

USA) [22] (Figure 1, step 1). Only one sample in data set 2a was discarded from further analysis, because it had too high a value for defective area percentage.

Individual data processing and analysis

Following quality control assessment, each data set was analyzed individually using the ArrayTrack™ tool (US Food and Drug Administration’s National Center for Toxicological Research, Jefferson, AR, USA) [23]. Array-Track is a comprehensive tool for microarray data storage, analysis, and interpretation that has been developed at the FDA’s National Center for Toxicological Research. To maintain consistency during the

Table 1 Information on data sets used for meta-analysis^a

Data set	Platform	Type of data	Samples, <i>n</i>	Source
Data set 1	1a: HG U133A 1b: HG U133B	Pediatric PBMC	59 (38 SLE, 21 controls)	Allantaz <i>et al.</i> [11]
Data set 2	2a: HG U95 Av1 2b: HG U95 Av2	Adult PBMC	90 (48 SLE, 42 controls)	Baechler <i>et al.</i> [12]
Data set 3	HG U133 Plus 2	Adult PBMC	58 (44 SLE, 14 controls)	Unpublished data
Data set 4	HG U95 Av1	Pediatric PBMC	39 (30 SLE, 9 controls)	Bennett <i>et al.</i> [35]
Data set 5	1a: HG U133A 1b: HG U133B	Pediatric PBMC	57 (47 SLE, 10 controls)	Chaussabel <i>et al.</i> [14]

^aPBMC, peripheral blood mononuclear cell; SLE, systemic lupus erythematosus.

individual analysis of data sets, similar normalization methods, statistical tests, and parameters were used with all data sets. First, all data sets except data set 4 were normalized using Robust Multi-array Analysis. Then Welch's *t* test was performed on each data set individually. The *P* value and fold change filters (0.01 and 1.5, respectively) were used to identify a unique list of DEGs from each data set (Figure 1, step 2). This list represented genes that were either notably upregulated or downregulated in the PBMCs of SLE patients when compared to the PBMCs of healthy controls. Each DEG list was then used to identify biological pathways significantly represented in SLE samples compared to the healthy controls ($P < 0.01$) in each data set (Figure 1, step 3). This pathway analysis was done using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems Inc., Redwood City, CA, USA).

Meta-analysis

Pathways common to all of the data sets were identified from the individual lists of pathways enriched in SLE patients compared to healthy controls (one for each data set) (Figure 1, step 4). The resulting list of pathways was indicative of processes significantly affected in all of the SLE data sets and comprised a pathway signature representative of all data sets and of the disease. From this common pathway signature, gene markers that met all of the following criteria were selected: (1) exhibited a fold change greater than 2 in at least one of the data sets (stringency increased from 1.5-fold to 2-fold to obtain a robust signature), (2) present in the DEG list in at least one of the data sets, and (3) involved in at least one of the commonly enriched pathways (Figure 1, step 5). These DEGs composed the collective signature (Figure 1, step 6).

Validation with the leave one data set out permutation method

To validate this technique, a leave one data set out permutation approach was employed (Figure 1, step 7). The meta-analysis technique described above was reiterated four times, each time leaving out one of the four data sets (data sets 1 to 4) and performing the analysis using the remaining three data sets. This gave rise to four different scenarios (Table 2). The gene signature obtained using the three data sets (for example, data sets 1 to 3) was then applied to the data set left out (for example,

data set 4). Unsupervised visualization techniques such as PCA and HCA were performed to examine how well the signature could differentiate SLE patients from healthy controls (Figures 2 and 3).

Metasignature development

Gene markers were generated for each of the four scenarios as described in the Meta-analysis section. Gene markers present in at least three of the four scenarios were grouped to comprise a 37-gene metasignature.

Confirmation using a fifth independent data set

Confirmation of the final 37-gene metasignature was done using an independent fifth data set, data set 5 (Figure 1, step 8). Again, PCA and HCA were carried out to evaluate the ability of the metasignature to differentiate SLE patients from healthy controls in this independent data set (Figure 4).

Results and discussion

Individual data sets of SLE and healthy control data sets derived from Affymetrix microarrays were analyzed using ArrayTrack following quality control (Figure 1, step 1) and normalization procedures. DEGs for individual data sets were identified using a *P* value cutoff of 0.01 and a fold change cutoff of 1.5 (Figure 1, step 2).

Biological pathways identified in SLE patients through the leave one data set out permutation method

After applying the leave one data set out approach for each of the four scenarios (Table 2), commonly enriched biological pathways were identified using IPA software (Table 3). Three biological pathways were consistently enriched in SLE patients in all four scenarios: interferon (IFN) signaling, interleukin (IL)-10 signaling, and glucocorticoid receptor signaling. An additional pathway, LXR/RXR signaling, was identified only in scenario IV.

Previous studies have provided evidence of increased autoimmunity in patients undergoing IFN treatment [24]. More specifically, there is evidence of women developing SLE during IFN- α treatment [25]. Several studies have shown upregulation of the IFN signaling pathway in SLE patients [9,12,17,26-30]. Therefore, it is understandable that IFN signaling appears to be affected across all data sets.

IL-10 signaling appears to be dysregulated and may be indicative of the inflammatory processes involved in

Table 2 Scenarios for leave one data set out validation

Scenario I	Scenario II	Scenario III	Scenario IV
Data set 1 (a and b)	Data set 2 (a and b)	Data set 1 (a and b)	Data set 1 (a and b)
Data set 2 (a and b)	Data set 3	Data set 3	Data set 2 (a and b)
Data set 3	Data set 4	Data set 4	Data set 4
Test: Data set 4	Test: Data set 1 (a and b)	Test: Data set 2 (a and b)	Test: Data set 3

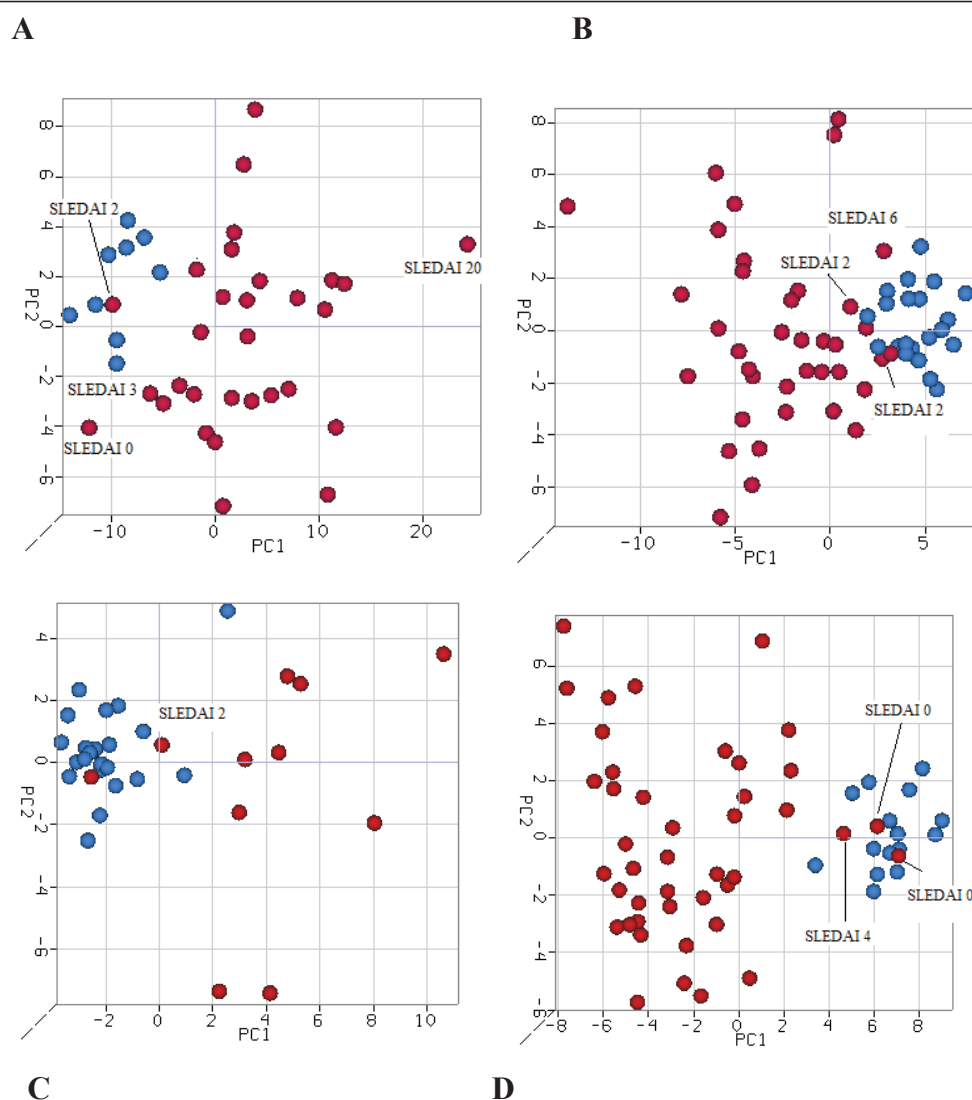


Figure 2 Principal component analysis from all scenarios. There is a clear distinction between healthy samples and systemic lupus erythematosus (SLE) patients, shown in blue and red, respectively. (A) Scenario I. (B) Scenario II. (C) Scenario III. (D) Scenario IV. The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) scores indicate that most SLE patients exhibiting close to normal gene expression are in remission (SLEDAI score 0) or have low disease activity (SLEDAI scores 2 and 3).

SLE. IL-10 binds to IL-10 receptor 1 on immune cells and activates the JAK-STAT signaling pathway, which is the key IFN signaling mechanism [31]. In support of the hypothesis that IL-10 is involved in SLE, IL-10 has been identified as one of risk loci for SLE in a large genome-wide association study [32].

Glucocorticoid receptors are also believed to influence cytokine signaling and may be indirectly involved in the pathways underpinning SLE [33]. In fact, glucocorticoids are routinely used in the treatment of SLE patients.

Genes differentially expressed in SLE

Each of the four scenarios produced a gene signature: scenario I produced a signature comprising 51 genes,

scenario II produced a signature with 31 genes, scenario III produced a signature with 34 genes, and scenario IV produced a signature with 28 genes. These DEGs represent the three main SLE disease pathways (IFN signaling, IL-10 signaling pathway, and glucocorticoid signaling pathways) as discussed in the section above.

A separate analysis of the pediatric and adult data sets used to identify DEGs and pathways in the two populations was conducted. Similar gene expression patterns were observed in adult and pediatric populations, although the extent of upregulation of some of the genes was higher in the pediatric data sets (unpublished results).

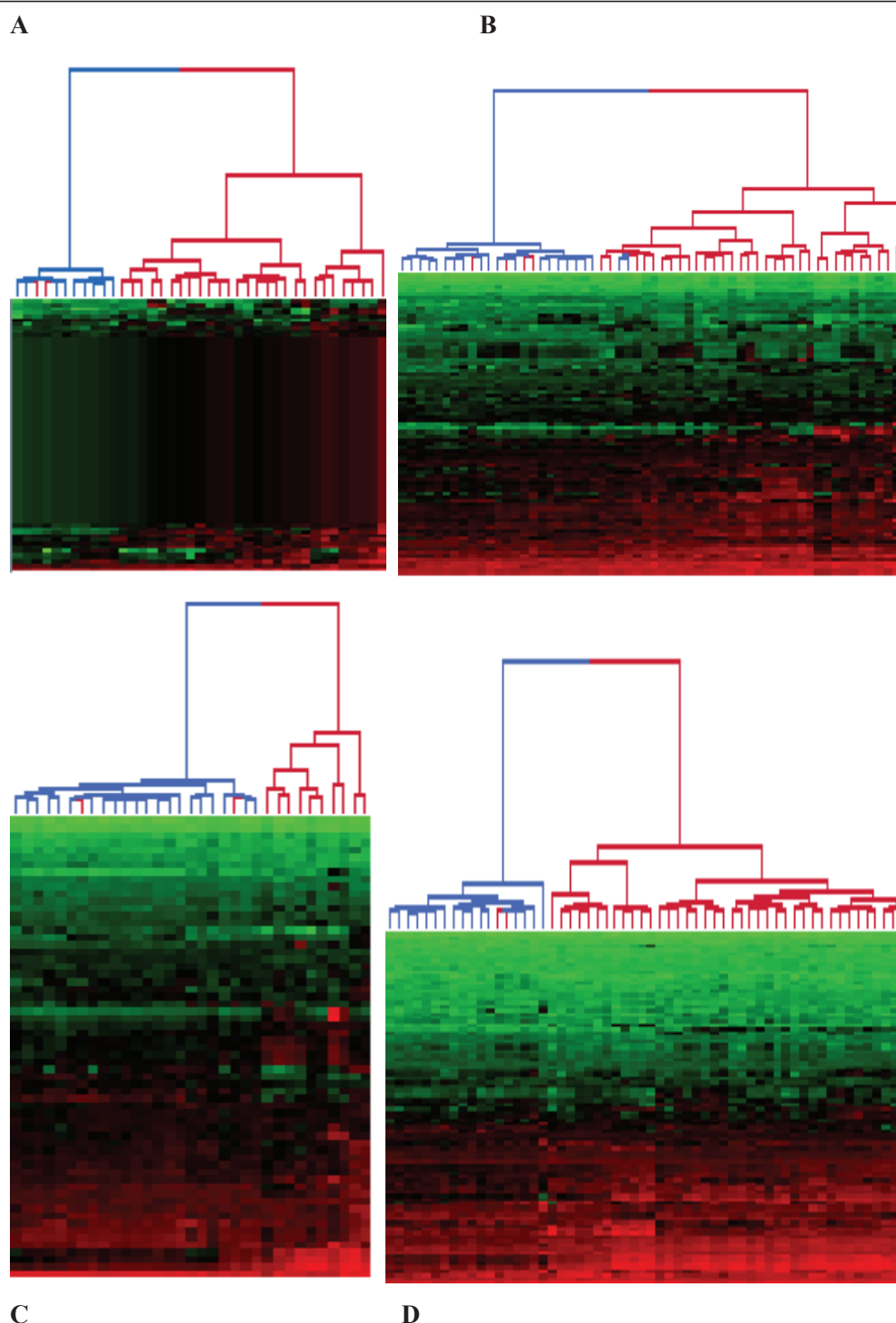
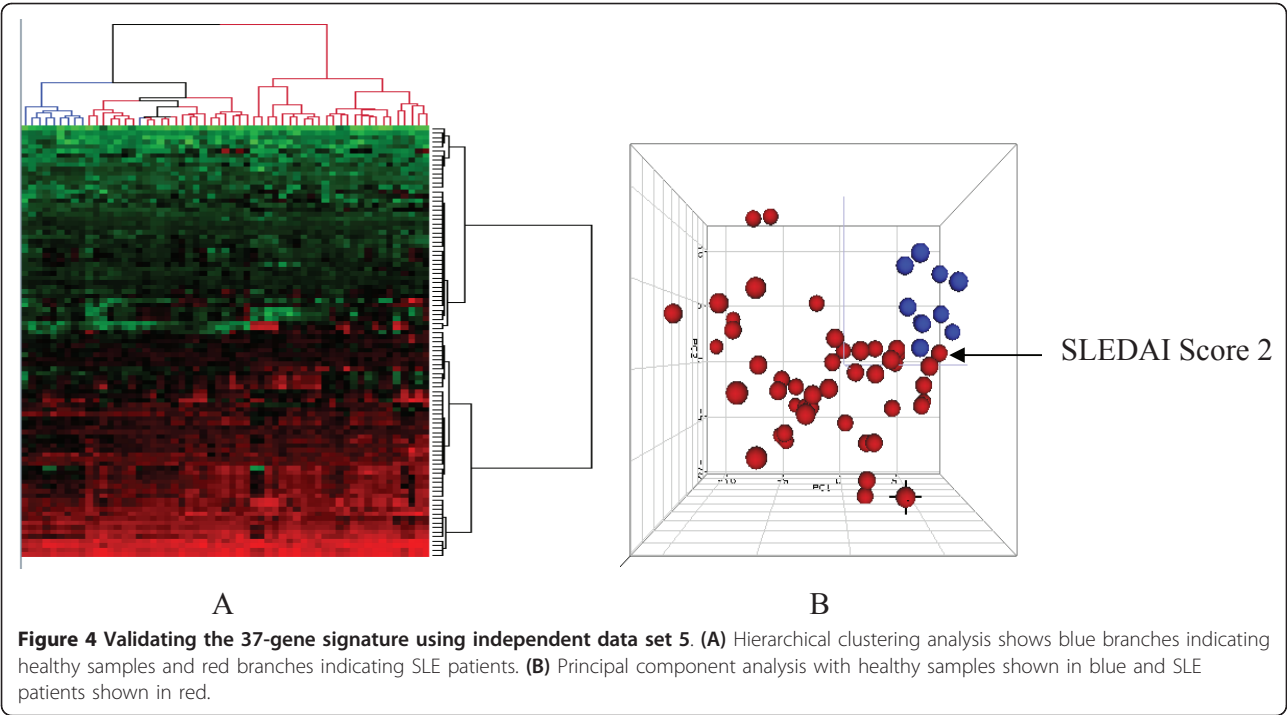


Figure 3 Hierarchical clustering analysis for all scenarios. Blue branches indicate healthy samples, and red branches indicate SLE patients. (A) Scenario I. (B) Scenario II. (C) Scenario III. (D) Scenario IV.

Validation of the meta-analysis approach

For each scenario, the signature obtained using the three data sets was applied to the fourth data set to observe how effectively the expression of the signature genes could distinguish between the SLE and healthy populations.

The PCAs and HCAs obtained for each scenario are presented in Figures 2 and 3, respectively. The PCA and HCA produced similar and consistent results. Grouping of samples based on the expression of signature genes alone produced a clear distinction between SLE patients



and healthy controls. The results suggest that the DEG signatures derived by using the leave one data set out permutation approach in the four scenarios (Table 2) can potentially identify a robust gene expression signature for SLE.

Gene expression signatures for SLE and Systemic Lupus Erythematosus Disease Activity Index scores

The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) is a validated scoring system that can be used to describe the range of disease activity and comprises a weighted score calculated by the presence or absence of 24 symptoms. The association of SLEDAI scores to expression profiles of SLE patients was evaluated. While the majority of the samples were grouped into their respective classes (SLE or control; see Figure 2), 12 SLE patients exhibited expression profiles similar to the control samples. On closer examination of these samples, it was found that the scores for nine of the patients indicated that they were either in remission

(SLEDAI score 0) or had mild activity of the disease (SLEDAI score 2 or 3). These findings lend further credence to the ability of the pathway-based meta-analysis approach used here in distinguishing SLE patients from healthy controls. Correlation between SLEDAI scores and gene expression signatures has also been reported in the literature [9,26,34].

Metasignature for SLE

A 37-gene signature was generated by the meta-analysis workflow (Table 4). Many IFN-induced genes involved in the IFN signaling pathway (Figure 5), such as *IFIT1*, *IFIT3*, *IFITM1*, *IFIT35*, *MX1*, and *OAS1*, were present in the signature. Overexpression of IFN-regulated genes in PBMCs of SLE patients has been reported in several publications [9,15,26,35–37]. In addition to genes involved in the IFN signaling pathway, genes in cytokine signal transduction (*SOCS1* and *SOCS3*) were also among the DEGs in SLE patients. Differential expression of many other biomarkers associated with inflammatory

Table 3 Biological pathways commonly and significantly enriched in the four scenarios^a

Scenario I	Scenario II	Scenario III	Scenario IV
IPA pathways	IPA pathways	IPA pathways	IPA pathways
IL-10 signaling	IL-10 signaling	IL-10 signaling	IL-10 signaling
Interferon signaling	Interferon signaling	Interferon signaling	Interferon signaling
Glucocorticoid receptor signaling	Glucocorticoid receptor signaling	Glucocorticoid receptor signaling	Glucocorticoid receptor signaling
			LXR/RXR signaling

^aIPA, Ingenuity Pathway Analysis software; IL, interleukin; LXR, liver X receptor; RXR, retinoid X receptor.

Table 4 Signature genes and their functionalities

Functionality	Genes from signature
Interferon signaling and interferon-induced or interferon-regulated genes	<i>IFI35, IFIT1, IFIT3, IFITM1, MX1, OAS1, STAT1, STAT2, CCL2, CXCL1, SOCS1, SOCS3</i>
Inflammatory and immune response	<i>CCL3, CCR1, CD163, FCGR1A, IL1R2, IL1B, IL1RN, IL-8, NFAT5, TRA@, MAP2K3, MAP2K6, SMAD3, FCGR2A, FCGR2B, NFKBIA</i>
Cellular proliferation and differentiation	<i>CDKN1C, CDK1A, DUSP1, EP300, FOS, JUN, PRKACA, PRKACB</i>
Protein folding	<i>SLP1</i>

and/or immune responses and with cellular proliferation was also observed, as shown in Table 4.

Confirmation of the metasignature using an independent data set

This signature was applied to an independent fifth data set (data set 5) to evaluate its ability to distinguish the SLE samples from the control samples. Figure 4 shows that the signature demonstrated clear differentiation

between SLE patients and healthy controls. In the HCA analysis, nine of ten healthy samples clustered together and were clearly separated from the cluster of SLE samples (Figure 4A). The PCA analysis also showed that the majority of the SLE samples and healthy samples were grouped separately (Figure 4B). The one SLE sample that was clustered with the healthy samples had a SLE-DAI score of 2, confirming our earlier observations with different data sets (Figure 2).

Conclusions

The novel pathway-based meta-analysis approach proved to be a useful technique for grouping disparate microarray data sets. This technique allowed for validated conclusions to be drawn across four different data sets and confirmed by testing on an independent fifth data set. Since the metasignature was obtained from pathways that were enriched in SLE samples across all of the data sets, it is highly representative of biological pathways related to SLE. The metasignature produced here may serve as a source for identifying potential

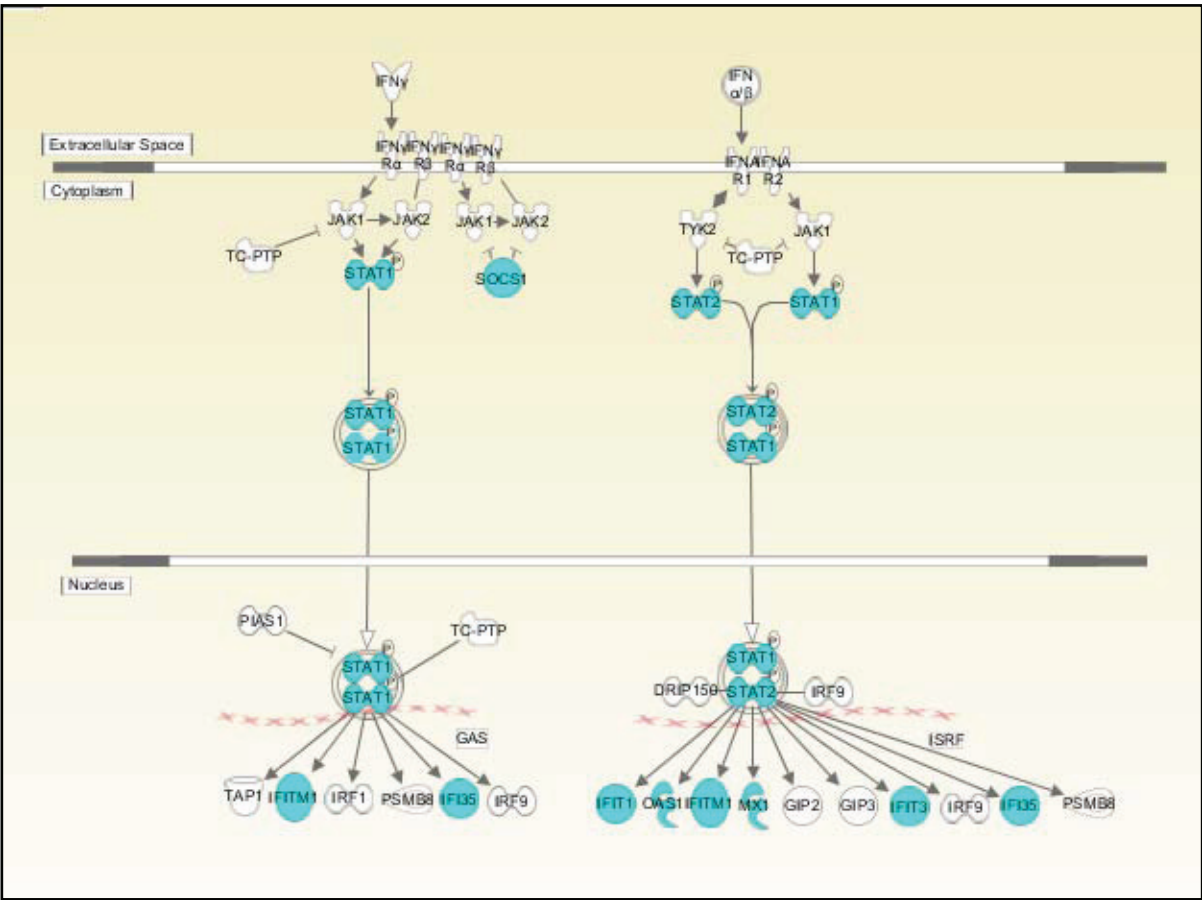


Figure 5 Interferon signaling pathways. Interferon-α, interferon-β, and interferon-γ signaling pathways are shown. The genes in blue represent differentially expressed genes that are part of the SLE metasignature.

therapeutic targets for SLE. Also, with further refinement, it might be made clinically more intuitive, which may also prove to be useful for the diagnosis and monitoring of SLE. Moreover, the meta-analysis approach outlined here provides a simple and intuitive solution for combining disparate microarray data sets to identify a robust metasignature.

Acknowledgements

The authors thank The Office of Women's Health, US Food and Drug Administration, for funding the research project (OWH study 07-09-0001-SA). The authors also thank all the prominent research groups in the SLE field who contributed data for the meta-analysis: Dr Thomas Aune, Dr Emily Baechler Gillespie, Dr Mary Crow, Dr Bernard Lauwerys, GeneLogic, Dr Dafna Gladman, Dr Virginia Pascual, and Dr Violeta Rus. Of all data sets, only data generated with oligonucleotide microarrays were selected for the meta-analysis described in this article. The views presented in this article do not necessarily reflect those of the US Food and Drug Administration.

Author details

¹National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA. ²Genomics Group, Office of Clinical Pharmacology, Office of Translational Science, Center for Drug Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA. ³Z-Tech Corporation, an ICF International Company at the National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, HFT 230, Jefferson, AR 72079, USA. ⁴University of Texas, Austin, Texas Institute for Drug and Diagnostic Development, 2500 Speedway, TX 78712, USA.

Authors' contributions

DA created the first draft of the manuscript and performed the meta-analysis. PM and SA performed the analysis to confirm the metasignature in an independent data set. SA, PM, and HF helped significantly to draft the manuscript. SA and WT helped coordinate the project and finalize the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 10 February 2011 Accepted: 30 May 2011

Published: 30 May 2011

References

- Santos ES, Perez CA, Raez LE: How is gene-expression profiling going to challenge the future management of lung cancer? *Future Oncol* 2009, **5**:827-835.
- Gómez-Raposo C, Mendiola M, Barriuso J, Hardisson D, Redondo A: Molecular characterization of ovarian cancer by gene-expression profiling. *Gynecol Oncol* 2010, **118**:88-92.
- Verweij CL: Transcript profiling towards personalised medicine in rheumatoid arthritis. *Neth J Med* 2009, **67**:364-371.
- Slodkowska EA, Ross JS: MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 2009, **9**:417-422.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* 2004, **101**:9309-9314.
- Choi H, Shen R, Chinnaiyan AM, Ghosh D: A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics* 2007, **8**:364.
- Wang J, Coombes KR, Highsmith WE, Keating MJ, Abruzzo LV: Differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells: a meta-analysis of three microarray studies. *Bioinformatics* 2004, **20**:3166-3178.
- Shen R, Ghosh D, Chinnaiyan AM: Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* 2004, **5**:94.
- Nikpour M, Dempsey AA, Urowitz MB, Gladman DD, Barnes DA: Association of a gene expression profile from whole blood with disease activity in systemic lupus erythematosus. *Ann Rheum Dis* 2008, **67**:1069-1075.
- Lyons PA, McKinney EF, Rayner TF, Hattton A, Woffendin HB, Koukoulaki M, Freeman TC, Jayne DR, Chaudhry AN, Smith KG: Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. *Ann Rheum Dis* 2010, **69**:1208-1213.
- Allantaz F, Chaussabel D, Stichweh D, Bennett L, Allman W, Mejias A, Ardura M, Chung W, Smith E, Wise C, Palucka K, Ramilo O, Punaro M, Banchereau J, Pascual V: Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. *J Exp Med* 2007, **204**:2131-2144.
- Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, Shark KB, Grande WJ, Hughes KM, Kapur V, Gregersen PK, Behrens TW: Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 2003, **100**:2610-2615.
- Pascual V, Banchereau J, Palucka AK: The central role of dendritic cells and interferon- α in SLE. *Curr Opin Rheumatol* 2003, **15**:548-556.
- Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H, Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V: A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008, **29**:150-164.
- Sawalha AH, Jeffries M, Webb R, Lu Q, Gorelik G, Ray D, Osban J, Knowlton N, Johnson K, Richardson B: Defective T-cell ERK signaling induces interferon-regulated gene expression and overexpression of methylation-sensitive genes similar to lupus patients. *Genes Immun* 2008, **9**:368-378.
- Teramoto K, Negoro N, Kitamoto K, Iwai T, Iwao H, Okamura M, Miura K: Microarray analysis of glomerular gene expression in murine lupus nephritis. *J Pharmacol Sci* 2008, **106**:56-67.
- Brenet F, Socci ND, Sonenberg N, Holland EC: Akt phosphorylation of La regulates specific mRNA translation in glial progenitors. *Oncogene* 2009, **28**:128-139.
- Petri M, Singh S, Tesfayone H, Dedrick R, Fry K, Lal P, Williams G, Bauer J, Gregersen P, Behrens T, Baechler E: Longitudinal expression of type I interferon responsive genes in systemic lupus erythematosus. *Lupus* 2009, **18**:980-989.
- Thibault DL, Graham KL, Lee LY, Balboni I, Hertzog PJ, Utz PJ: Type I interferon receptor controls B-cell expression of nucleic acid-sensing Toll-like receptors and autoantibody production in a murine model of lupus. *Arthritis Res Ther* 2009, **11**:R112.
- Liu K, Li QZ, Delgado-Vega AM, Abelson AK, Sánchez E, Kelly JA, Li L, Liu Y, Zhou J, Yan M, Ye Q, Liu S, Xie C, Zhou XJ, Chung SA, Pons-Estel B, Witte T, de Ramón E, Bae SC, Barizzone N, Sebastiani GD, Merrill JT, Gregersen PK, Gilkeson GG, Kimberly RP, Vyse TJ, Kim I, D'Alfonso S, Martin J, Harley JB, Criswell LA, Profile Study Group; Italian Collaborative Group; German Collaborative Group; Spanish Collaborative Group; Argentinian Collaborative Group; SLEGEN Consortium, Wakeland EK, Alarcón-Riquelme ME, Mohan C: Kallikrein genes are associated with lupus and glomerular basement membrane-specific antibody-induced nephritis in mice and humans. *J Clin Invest* 2009, **119**:911-923.
- Silva GL, Junta CM, Mello SS, Garcia PS, Rassi DM, Sakamoto-Hojo ET, Donadi EA, Passos GA: Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. *Ann N Y Acad Sci* 2007, **1110**:33-46.
- GeneData Expressionist. 2007 [http://www.genedata.com/resources/press-release-archive/press-releases/article/genedata-provides-expressionist-platform-to-us-food-and-drug-administration.html?tx_ttnews%5Bp5%5D=1245776672&tx_ttnews%5BbackPid%5D=82&cHash=aa431e281f].
- Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, Harris A, Hong H, Xie Q, Perkins R, Shi L, Casciano D: ArrayTrack: supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ Health Perspect* 2003, **111**:1819-1826.

24. Ronnblom L, Eloranta ML, Alm GV: **The type I interferon system in systemic lupus erythematosus.** *Arthritis Rheum* 2006, **54**:408-420.
25. Niewold TB, Swedler WI: **Systemic lupus erythematosus arising during interferon-α therapy for cryoglobulinemic vasculitis associated with hepatitis C.** *Clin Rheumatol* 2005, **24**:178-181.
26. Kirou KA, Lee C, George S, Louca K, Peterson MG, Crow MK: **Activation of the interferon-α pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease.** *Arthritis Rheum* 2005, **52**:1491-1503.
27. Ronnblom L: **Potential role of IFNα in adult lupus.** *Arthritis Res Ther* 2010, **12**(Suppl 1):S3.
28. Wang S, Yang N, Zhang L, Huang B, Tan H, Liang Y, Li Y, Yu X: **Jak/STAT signaling is involved in the inflammatory infiltration of the kidneys in MRL/lpr mice.** *Lupus* 2010, **19**:1171-1180.
29. Perl A: **Emerging new pathways of pathogenesis and targets for treatment in systemic lupus erythematosus and Sjögren's syndrome.** *Curr Opin Rheumatol* 2009, **21**:443-447.
30. Pascual V, Farkas L, Banchereau J: **Systemic lupus erythematosus: all roads lead to type I interferons.** *Curr Opin Immunol* 2006, **18**:676-682.
31. Hermann J, Gruber S, Neufeld JB, Grundtner P, Graninger M, Graninger WB, Berghold A, Gasche C: **IL10R1 loss-of-function alleles in rheumatoid arthritis and systemic lupus erythematosus.** *Clin Exp Rheumatol* 2009, **27**:603-608.
32. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, et al: **A large-scale replication study identifies *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus.** *Nat Genet* 2009, **41**:1228-1233.
33. Rogatsky I, Ivashkiv LB: **Glucocorticoid modulation of cytokine signaling.** *Tissue Antigens* 2006, **68**:1-12.
34. Feng X, Wu H, Grossman JM, Hanvivadhanakul P, Fitzgerald JD, Park GS, Dong X, Chen W, Kim MH, Weng HH, Furst DE, Gorn A, McMahon M, Taylor M, Brahn E, Hahn BH, Tsao BP: **Association of increased interferon-inducible gene expression with disease activity and lupus nephritis in patients with systemic lupus erythematosus.** *Arthritis Rheum* 2006, **54**:2951-2962.
35. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, Pascual V: **Interferon and granulopoiesis signatures in systemic lupus erythematosus blood.** *J Exp Med* 2003, **197**:711-723.
36. Pascual V, Allantaz F, Patel P, Palucka AK, Chaussabel D, Banchereau J: **How the study of children with rheumatic diseases identified interferon-α and interleukin-1 as novel therapeutic targets.** *Immunol Rev* 2008, **223**:39-59.
37. Bauer JW, Petri M, Batiwalla FM, Koeuth T, Wilson J, Slattery C, Panoskaltis-Mortari A, Gregersen PK, Behrens TW, Baechler EC: **Interferon-regulated chemokines as biomarkers of systemic lupus erythematosus disease activity: a validation study.** *Arthritis Rheum* 2009, **60**:3098-3107.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1741-7015/9/65/prepub

doi:10.1186/1741-7015-9-65

Cite this article as: Arasappan et al.: Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. *BMC Medicine* 2011 **9**:65.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

